AD-A077 458   SOUTH CAROLINA UNIV COLUMBIA DEPT OF MATHEMATICS COM--ETC F/G 12/1
                A BAYESIAN NONPARAMETRIC ESTIMATOR OF SURVIVAL PROBABILITY ASSU--ETC(U)
                AUG 79   W J PADGETT , L J WEI   F49620-79-C-0140
UNCLASSIFIED   TR-42   AFOSR-TR-79-1033   NL

|OF|
AD-
A077458

END
DATE
FILMED
12-79
DDC

See 1473
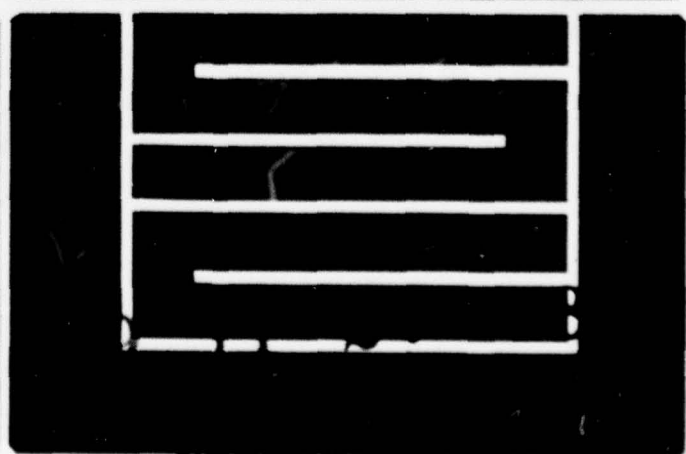
LEVEL

D D C
RECEIVED
NOV 30 1979

A

ADA077458

JDC FILE COPY

A BAYESIAN NONPARAMETRIC ESTIMATOR OF SURVIVAL
PROBABILITY ASSUMING INCREASING FAILURE RATE*

by

W. J. Padgett and L. J. Wei
University of South Carolina
Statistics Technical Report No. 42
62C10

distribution unlimited.
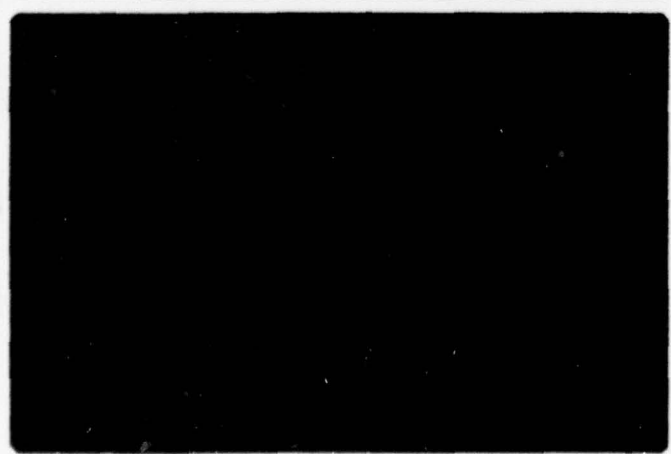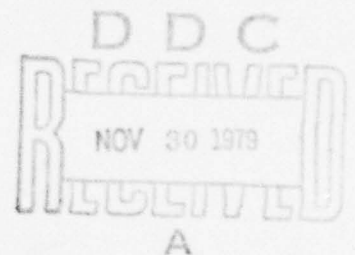
A BAYESIAN NONPARAMETRIC ESTIMATOR OF SURVIVAL

PROBABILITY ASSUMING INCREASING FAILURE RATE[*]

by

W. J. Padgett and L. J. Wei
University of South Carolina
Statistics Technical Report No. 42
62C10

August, 1979

Department of Mathematics, Computer Science, and Statistics
University of South Carolina
Columbia, South Carolina 29208

Abstract

Bayesian nonparametric estimators of the survival function, the failure rate function, and the density function are obtained using jump processes as prior distributions on the space of increasing failure rate functions. The jump processes are more appealing intuitively than previously used Dirichlet processes, integral gamma process, and processes neutral to the right and have a meaningful physical interpretation. Examples are given and the estimates are compared with the maximum likelihood estimates. In addition, the Bayesian nonparametric estimators are presented for arbitrarily right-censored observations.

Accession For

NTIS GRA&I
DDC TAB   -
Unannounced
Justification

By

Availability Codes
Avail and/or
Special

A

1. INTRODUCTION

The problem of Bayesian nonparametric estimation of a survival or reliabil-
ity function has been considered in recent years by several authors (for
example, see Ferguson (1973, 1974), Ferguson and Phadia (1979), Lo (1979),
and Susarla and Van Ryzin (1976)). These authors have utilized mainly the
Dirichlet processes, integral gamma processes, or processes neutral to the
right as prior probability distributions over the space of all distribution
functions of interest. The Dirichlet process priors put all of the probability
on discrete distributions while processes neutral to the right and integral
gamma processes are somewhat difficult to use in practice. Also, if further
prior information concerning the distribution function F to be estimated is
known, the Dirichlet process priors may not be entirely appropriate. For
example, if it is assumed that F has an increasing failure rate (IFR)
function, the Dirichlet process would not be suitable since F is absolutely
continuous on at least part of its support. In this paper we propose a prior
process on the failure rate function for Bayes estimation of the failure rate
and survival functions under the assumption of IFR. The proposed prior is
very practical and has a physical interpretation similar to that of a shock
model with shocks occurring as a Poisson process. No assumptions are made
about F except that it is IFR.

Let $r(t)$ denote the true failure rate function of the unknown distribu-
tion function F and let $\bar{F}(t)$ denote survival probability at $t \geq 0$. We
assume that F has a density function $f(t) = r(t)\bar{F}(t)$. Nonparametric
estimation of the failure rate function has been considered by Watson and
Leadbetter (1964a, 1964b) and by Barlow and van Zwet (1971), among others.
In particular, Grenander (1956) and Marshall and Proschan (1965) obtained the

maximum likelihood estimator of $r(t)$ under increasing assumptions. Their estimator was a step function and took the value $+\infty$ at the largest value of the sample. Padgett and Wei (1979) have recently obtained the maximum likelihood estimator of monotone failure rate functions for arbitrarily right-censored survival data.

The prior probability distribution on $r(t)$ which we propose is motivated by the fact that $r(t)$ can be closely approximated and estimated by a step function. We assume throughout the paper that $r(t)$ is <u>increasing</u> (meaning nondecreasing) for $t > 0$ so F is an IFR life distribution. Suppose the experimenter has prior information that he can quantify as a jump process for $r(t)$ with constant jumps of size ϵ at times $T_i > 0$, $i = 1, 2, \cdots$, where T_i are arrival times for a Poisson process $\{N(t) : t \geq 0\}$ with intensity rate ν. This constant jump process thus describes the prior probability distribution over the parameter space $\Theta = \{r : \quad r$ is an increasing failure rate function$\}$. This prior distribution over Θ also has the following intuitively appealing physical interpretation: Suppose shocks occur randomly in time to a system or component causing a certain amount of damage to it which increases the failure rate of such components by a certain constant positive amount ϵ. Assume that the damage is cumulative and that the shocks occur according to a Poisson process $\{N(t) : t \geq 0\}$ with intensity ν. Then the failure rate function can be considered, at least <u>a priori</u>, as a constant jump process. This physical interpretation is similar to those probability models proposed by Gaver (1963), Antelman and Savage (1965), and Reynolds and Savage (1971) in which the failure rate function was stochastic.

In Section 2 we will utilize the constant jump process as a prior distribution on increasing failure rate functions to obtain a Bayes nonparametric

estimator of $r(t)$ and of the reliability or survival function $\bar{F}(t) = \exp\{-\int_0^t r(u)du\}$ based on a random sample of size n from $F(t)$. This estimator of $r(t)$ turns out to be a smooth curve rather than a step function. In Section 3, an example is given using data simulated by Dubey (1967) from a Weibull distribution with increasing failure rate. Finally, in Section 4, it is shown that the same techniques may be easily applied to obtain Bayesian nonparametric estimators of $r(t)$ and $\bar{F}(t)$ for arbitrarily right-censored data. This problem was studied by Susarla and Van Ryzin (1976) for Dirichlet process priors and by Ferguson and Phadia (1979) for processes neutral to the right.

2. THE BAYESIAN NONPARAMETRIC ESTIMATORS FOR COMPLETE SAMPLES

Let $R(t)$ denote the failure rate function with prior distribution given by the constant jump process described in Section 1. Then from the jump size $\epsilon > 0$ and Poisson process $\{N(t) : t \geq 0\}$, we have $R(t) = \epsilon N(t)$ with the corresponding survival function given by

$$\bar{F}(t) = \exp[-\int_0^t R(u)du] = \exp[-\epsilon \sum_{i=1}^{N(t)} (t-T_i)], \quad t \geq 0, \tag{2.1}$$

where T_i, $i = 1, 2, \cdots, N(t)$, denote the "arrival" times of jumps of size ϵ in $(0,t]$. Given $N(t) = k$ and t, it is well-known that $T_1, \cdots, T_k$ are independent uniformly distributed random variables on $(0,t]$.

Let $x_1, \cdots, x_n$ be a given sample of size n from the unknown distribution function F. It is desired to obtain Bayes estimators $\hat{\bar{F}}(t)$ and $\hat{R}(t)$ of $\bar{F}(t)$ and $r(t)$, respectively, with a squared-error loss function of the form $L(\bar{F}, \hat{\bar{F}}) = \int_{-\infty}^{\infty} [\bar{F}(t) - \hat{\bar{F}}(t)]^2 dw(t)$, where $w(t)$ is an arbitrary non-random weight function.

The likelihood function of the sample is

$$\prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n} R(x_i) \exp[-\int_0^{x_i} R(u)du]$$

$$= \prod_{i=1}^{n} \{\epsilon N(x_i) \exp[-\epsilon \sum_{j=1}^{N(x_i)} (x_i - T_j)]\} . \tag{2.2}$$

Thus, at least theoretically, the posterior probability measure P_n over the parameter space $\Theta = \{r : r$ is an increasing failure rate function$\}$ may be obtained from

$$P_n(B) = [\int_B \prod_{i=1}^{n} f(x_i)dP]/[\int_\Theta \prod_{i=1}^{n} f(x_i)dP] ,$$

where B is a measurable subset of Θ and P denotes the prior probability measure over Θ induced by the constant jump process. It is not necessary, however, to find P_n in order to obtain the desired Bayes estimators. It is much simpler to calculate the posterior mean of $\overline{F}(t)$ or $R(t)$ directly. We first obtain the Bayes estimate of $\overline{F}(t)$ from the expression

$$\hat{\overline{F}}(t) = E[\overline{F}(t)|x_1, \cdots, x_n]$$

$$= [\int_\Theta \overline{F}(t) \prod_{i=1}^{n} f(x_i)dP]/[\int_\Theta \prod_{i=1}^{n} f(x_i)dP] . \tag{2.3}$$

Assume without any loss of generality that the sample values are increasing. $x_1 \leq x_2 \leq \cdots \leq x_n$. Then from equation (2.2), the denominator of (2.3) becomes

$$\int_\Theta \prod_{i=1}^{n} f(x_i)dP = \int \prod_{i=1}^{n} \{\epsilon N(x_i)\exp[-\epsilon \sum_{j=1}^{N(x_i)} (x_i - T_j)]$$

$$\times dP(N(x_1), \cdots, N(x_n), T_1, \cdots, T_{N(x_n)})$$

$$= E_{N(x_1), \cdots, N(x_n)} \{ E_{\underline{T}} | N(x_i) = k_i, i=1, \cdots, n [\epsilon \prod_{i=1}^{n} k_i \exp(-\epsilon \sum_{i=1}^{n} \sum_{j=1}^{k_i} (x_i - T_j))] \} ,$$

$$(2.4)$$

where $\underline{T} = (T_1, \cdots, T_{N(x_n)})$. The exponent in (2.4) may be rewritten in the form

$$-\epsilon \sum_{i=1}^{n} \sum_{j=1}^{k_i} (x_i - T_j) = -\epsilon [k_1 \sum_{i=1}^{n} x_i + (k_2 - k_1) \sum_{i=1}^{n} x_i + \cdots + (k_n - k_{n-1}) x_n]$$

$$+ \epsilon [n \sum_{j=1}^{k_1} T_j + (n-1) \sum_{j=k_1+1}^{k_2} + \cdots + \sum_{j=k_{n-1}+1}^{k_n} T_j] .$$

Now, conditioned on k_{i-1} jumps in the interval $(0, x_{i-1}]$, $T_{k_{i-1}+1}, \cdots, T_{k_i}$ are order statistics from independent uniformly distributed random variables on $(x_{i-1}, x_i]$, $i = 2, \cdots, n$ and $T_1, \cdots, T_{k_1}$ are order statistics from independent uniformly distributed random variables on $(0, x_1]$. Hence, letting $x_0 = 0$ and $k_0 = 0$, we have for the ith interval $(x_{i-1}, x_i]$, $i = 1, \cdots, n$,

$$\int_{x_{i-1}}^{x_i} \cdots \int_{x_{i-1}}^{x_i} \exp[(n-i+1) \epsilon \sum_{j=k_{i-1}+1}^{k_i} t_j] \prod_{j=k_{i-1}+1}^{k_i} dt_j / (x_i - x_{i-1})$$

$$(2.5)$$

$$= \{ [\exp(\epsilon(n-i+1)x_i) - \exp(\epsilon(n-i+1)x_{i-1})] / [\epsilon(n-i+1)(x_i - x_{i-1})] \}^{k_i - k_{i-1}} .$$

By the mean value theorem, there is an $x_i^0 \epsilon (x_{i-1}, x_i]$ for which

$$\frac{\exp[(n-i+1) \epsilon x_i] - \exp[(n-i+1) \epsilon x_{i-1}]}{(n-i+1)(x_i - x_{i-1})} = \epsilon \exp[(n-i+1) \epsilon x_i^0] ,$$

$$(2.6)$$

so (2.5) can be simplified to $\exp[\epsilon (n-i+1) x_i^0 (k_i - k_{i-1})]$. Thus, summing over the appropriate Poisson variables, the expectation in the right-hand side of

(2.4) becomes

$$\epsilon^n e^{-\nu x_n} \sum_{k_n=0}^{\infty} \sum_{k_{n-1}=0}^{k_n} \cdots \sum_{k_1=0}^{k_2} (\prod_{i=1}^{n} k_i) \exp\{\epsilon \sum_{j=1}^{n} [(n-j+1)x_j^0]$$

$$- \sum_{i=j}^{n} x_i](k_j - k_{j-1})\} \nu^{k_n} \prod_{i=1}^{n} (x_i - x_{i-1})^{k_i - k_{i-1}} / \prod_{i=1}^{n} (k_i - k_{i-1})! \quad . \quad (2.7)$$

Equation (2.6) can be solved for

$$x_i^0 = \{\ln[\exp((n-i+1)\epsilon x_i) - \exp((n-i+1)\epsilon x_{i-1})]$$

$$- \ln[(n-i+1)\epsilon(x_i - x_{i-1})]\}/\epsilon(n-i+1)$$

which can be substituted into (2.7) to obtain the exact expression for the denominator.

To calculate the numerator of (2.3) the technique is the same as for the denominator except that we have an extra term. The numerator is

$$\int_\Theta \overline{F}(t) \prod_{i=1}^{n} f(x_i) dP$$

$$= E_{N(x_i), \cdots, N(x_n), N(t)} \{E_{\underline{T}|N(x_1)=k_1, \cdots, N(t)=k_t} [\exp[-\epsilon \sum_{j=1}^{k_t} (t-T_j)]$$

$$\times \epsilon^n \prod_{i=1}^{n} k_i \exp[-\epsilon \sum_{i=1}^{n} \sum_{j=1}^{k_i} (x_i - T_j)]]\} \quad . \quad (2.8)$$

Note that the order of summation corresponding to the first expectation in (2.8) depends on the value of t. If $x_{i-1} < t \le x_i$, $i = 1, \cdots, n+1$, where $x_0 = 0$ and $x_{n+1} = \infty$, then the value of $N(t)$ is between $N(x_{i-1})$ and $N(x_i)$ with probability one, with $N(0) = 0$ and $N(\infty) = \infty$. Hence,

we can simply relabel the x_i's and t as $n+1$ new points by defining

$$y_j = y_j(t) = \begin{cases} x_j, & j \le i-1 \\ t, & j = i \\ x_{j-1}, & j \ge i+1 \end{cases}, \quad i = 1, 2, \cdots, n. \tag{2.9}$$

Then similar to the denominator, the right-hand side of (2.8) becomes

$$E_{N(y_1), \cdots, N(y_{n+1})} \{ E_{\underline{T} | N(y_1) = k_1, \ldots, N(y_{n+1}) = k_{n+1}} [\epsilon^n \prod_{\substack{j=1 \\ j \ne i}}^{n+1} k_j$$

$$\times \exp[-\epsilon \sum_{j=1}^{n+1} \sum_{m=1}^{k_j} (y_j - T_m)]]\}$$

$$= \epsilon^n \exp(-\nu y_{n+1}) \sum_{k_{n+1}=0}^{\infty} \sum_{k_n=0}^{k_{n+1}} \cdots \sum_{k_1=0}^{k_2} (\prod_{\substack{j=1 \\ j \ne i}}^{n+1} k_j) \nu^{k_{n+1}}$$

$$\times \prod_{j=1}^{n+1} (y_j - y_{j-1})^{k_j - k_{j-1}} \exp[\epsilon \sum_{j=1}^{n+1} ((n-j+2)y_j^0 - \sum_{m=j}^{n+1} y_m)(k_j - k_{j-1})]$$

$$/ \prod_{j=1}^{n+1} (k_j - k_{j-1})!, \qquad x_{i-1} < t \le x_i, \quad i = 1, \cdots, n+1. \tag{2.10}$$

Therefore, the desired Bayes estimator $\hat{\bar{F}}(t)$ of survival probability $\bar{F}(t)$ is given by (2.10) divided by (2.7). This estimator is readily computable by electronic computer as the example in Section 3 shows.

Now, for the Bayesian nonparametric estimator of $r(t)$, the failure rate function, we simply calculate the required numerator expression for the posterior mean of $R(t) = \epsilon N(t)$, which is

$$\int_\Theta R(t) \prod_{i=1}^n f(x_i) dP = \int_\Theta R(t) \prod_{i=1}^n R(x_i) \exp[-\int_0^{x_i} R(u)du]dP$$

$$= \int \epsilon N(t) \prod_{i=1}^n [\epsilon N(x_i) \exp(-\epsilon \sum_{j=1}^{N(x_i)} (x_i - T_j))]dP(N(x_1), \cdots, N(t), \underline{T})$$

$$= E[\epsilon^{n+1}(\prod_{j=1}^n N(x_i))N(t) \exp(-\epsilon \sum_{i=1}^n \sum_{j=1}^{N(x_i)} (x_i - T_j))] . \qquad (2.11)$$

This expression is analogous to (2.8). Hence, relabeling the x_i's and t as y_j's as in (2.9), it is easily shown that (2.11) is equal to

$$E_{N(y_1), \cdots, N(y_{n+1})}\{E_{\underline{T}|N(y_1)=k_1, \cdots, N(y_{n+1})=k_{n+1}}[\epsilon^{n+1} (\prod_{j=1}^{n+1} k_j)$$

$$\times \exp[-\epsilon \sum_{j=1, j\neq i}^{n+1} \sum_{m=1}^{k_j} (y_j - T_m)]]\}$$

$$= \epsilon^{n+1} \exp(-\nu y_{n+1}) \sum_{k_{n+1}=0}^\infty \sum_{k_n=0}^{k_{n+1}} \cdots \sum_{k_1=0}^{k_2} (\prod_{j=1}^{n+1} k_j) \nu^{k_{n+1}}$$

$$\times \prod_{j=1}^{n+1} (y_j - y_{j-1})^{k_j - k_{j-1}} \exp[\epsilon \sum_{j=1}^{i-1} ((n-j+1)y_j^0 - \sum_{\substack{m=j \\ m\neq i}}^{n+1} y_m)$$

$$\times (k_j - k_{j-1}) + \epsilon \sum_{j=i+1}^{n+1} ((n-j+2)y_j^0 - \sum_{m=j}^{n+1} y_m)(k_j - k'_{j-1})]$$

$$/ \sum_{j=1}^{n+1} (k_j - k_{j-1})! , \qquad x_{i-1} < t \leq x_i , \quad i = 1, \cdots, n+1 , \qquad (2.12)$$

where $y_{i+1}^0 \varepsilon (y_{i-1}, y_{i+1}]$ as in (2.6), $k_j' = k_{i-1}$ for $j \le i$, and $k_j' = k_j$ for $j > i$.

Therefore, the Bayes estimator $\hat{R}(t)$ of $r(t)$ is given by (2.12) divided by (2.7). This estimator is a smooth curve, rather than a step function as the estimator of Marshall and Proschan (1965) yields, and is an increasing function over the range of the data so that we have a "closure" property of IFR over this range.

We remark that the constant jump size ε may be replaced by a random jump size W, say, with distribution function G. This extension complicates the expressions and makes them intractable. The constant jump prior process seems reasonable and sufficient for Bayesian analysis as was discussed in Section 1.

3. AN EXAMPLE

As an example, the formulas for the Bayesian nonparametric estimators derived in Section 2 were programmed for computation on an electronic computer. A random sample of size $n = 5$ was selected from the 100 observations of Dubey (1967) generated from a Weibull distribution with scale parameter one and shape parameter 1.2, so that the failure rate function is increasing. The sample selected given in order is 0.135873, 0.666654, 0.948871, 1.341265, 1.521437. Figures 1 and 2 show the resulting estimates of $r(t)$ and $\bar{F}(t)$, respectively. Several values of the prior process parameters ν and ε were used to illustrate the effects of the choices. Also, computed and plotted in the figures were the maximum likelihood estimates (Marshall and Proschan (1965)) and the true $r(t)$ and $\bar{F}(t)$ for comparison. The mle of $r(t)$ is $+ \infty$ at 1.521437 and beyond.
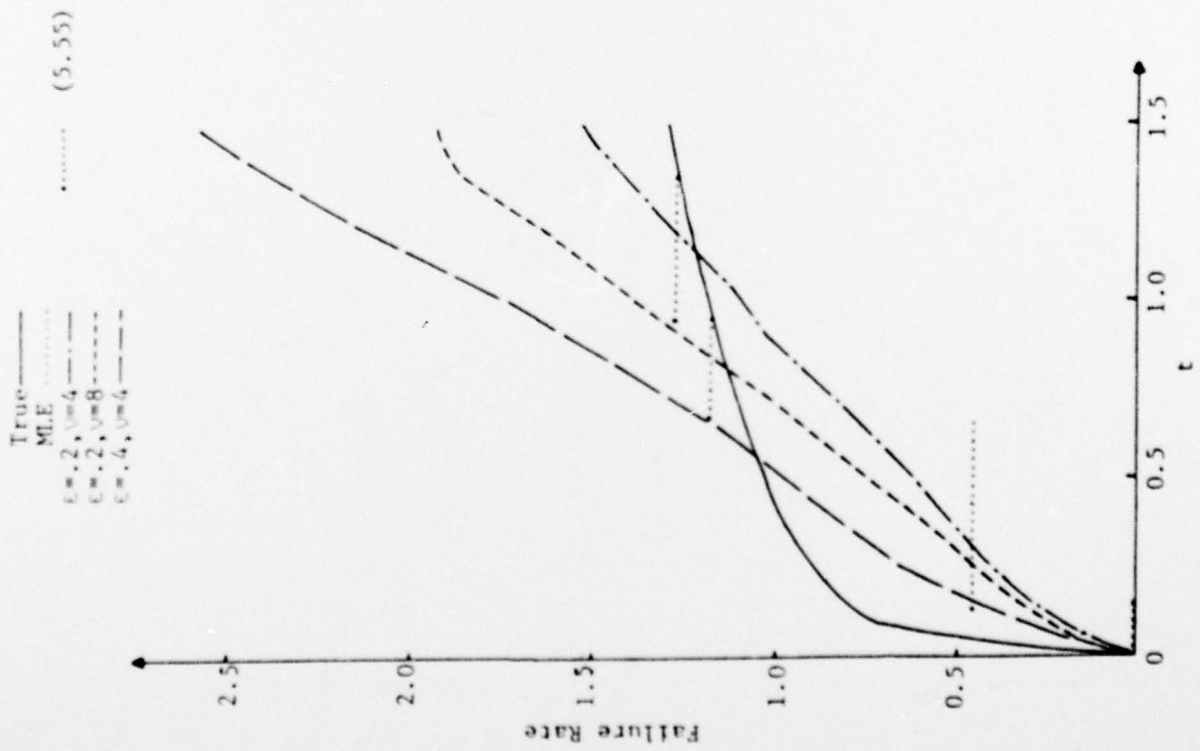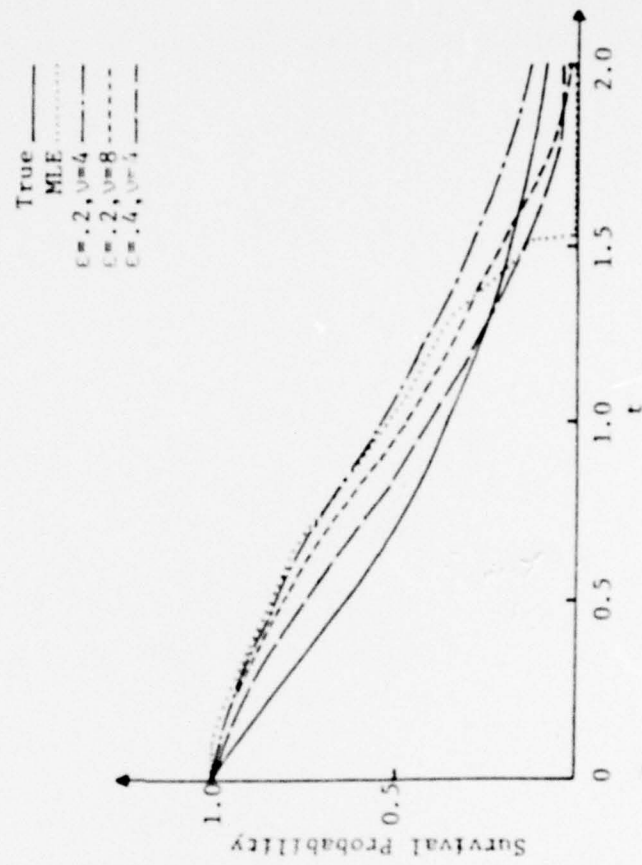
Figure 2. Estimates of Survival Function.

Figure 1. Estimates of Failure Rate Function.

The computations in Figures 1 and 2 were performed with 13 terms in the outside infinite summations in expressions (2.7), (2.10), and (2.12). The results were the same to at least two decimal places when using ten or nineteen terms. Also, an approximation to the x_i^0 in (2.6) (or y_j^0 in the numerator formulas) given by the midpoint $x_i^0 \approx \frac{x_i + x_{i-1}}{2}$ yielded the same results to at least two decimal places.

4. THE CASE OF ARBITRARILY RIGHT-CENSORED DATA

In many situations in life testing and survival analysis the items may be entered into and removed from the study at arbitrary times. Susarla and Van Ryzin (1976) and Ferguson and Phadia (1979) have obtained Bayesian non-parametric estimators of survival probability utilizing the Dirichlet processes and processes neutral to the right as priors, respectively.

Let $X_1^s, \cdots, X_n^s$ denote a random sample of true survival times of n items whose lifetime distribution is F. Let $U_1, \cdots, U_n$ be a set of constants or independent random variables which are also independent of $X_1^s, \cdots, X_n^s$. Define

$$X_i = \min\{X_i^s, U_i\}, \qquad \delta_i = \begin{cases} 1 & \text{if } X_i^s \le U_i \\ 0 & \text{if } X_i^s > U_i, \quad i = 1, \cdots, n. \end{cases}$$

The pairs (X_i, δ_i), $i = 1, \cdots, n$, are the observations which are available to the experimenter, in which X_i is an observed lifetime if $\delta_i = 1$ and a censored lifetime if $\delta_i = 0$. That is, it is known which observations represent "failures" and which ones represent "losses".

The likelihood function can be written as (Lagakos (1979))

$$L = \prod_{i=1}^{n} [f(x_i)]^{\delta_i} [\overline{F}(x_i)]^{1-\delta_i}$$

$$= \prod_{i=1}^{n} [R(x_i)]^{\delta_i} \overline{F}(x_i) \ ,$$

where $R(t)$ is the failure rate function as before. Hence, for the right-censored sample (X_i, δ_i), $i = 1, \cdots, n$, the Bayesian nonparametric estimator of $\overline{F}(t)$ is obtained as in Section 2 for complete samples. The estimator is given by

$$\hat{\overline{F}}(t) = E[\overline{F}(t) | (x_i, \delta_i), i=1, \cdots, n]$$

$$= \frac{\int_{\Theta} \overline{F}(t) \prod_{i=1}^{n} [R(x_i)]^{\delta_i} \exp[-\int_0^{x_i} R(u)du\,]dP}{\int_{\Theta} \prod_{i=1}^{n} [R(x_i)]^{\delta_i} \exp[-\int_0^{x_i} R(u)du]dP} \ , \qquad (4.1)$$

where the notation is the same as before, with $x_1 \le \cdots \le x_n$.

The denominator of (4.1) is given by

$$\exp(-\nu x_n) \sum_{k_n=0}^{\infty} \sum_{k_{n-1}=0}^{k_n} \cdots \sum_{k_1=0}^{k_2} \prod_{i=1}^{n} (\epsilon k_i)^{\delta_i} \nu^{k_n} \prod_{i=1}^{n} (x_i - x_{i-1})^{k_i - k_{i-1}}$$

$$\times \exp[\epsilon \sum_{j=1}^{n} ((n-j+1)x_j^0 - \sum_{i=j}^{n} x_i)(k_j - k_{j-1})] / \prod_{i=1}^{n} (k_i - k_{i-1})! \ . \qquad (4.2)$$

For the numerator of (4.1), we relabel the observations in a manner similar to that in Section 2. When $x_{i-1} < t \le x_i$, $i = 1, \cdots, n+1$, with $x_0 = 0$ and $x_{n+1} = \infty$, let

$$y_j = y_j(t) = \begin{cases} x_j & , \quad j \le i-1 \\ t & , \quad j = i \\ x_{j-1} & , \quad j \ge i+1 \end{cases} \quad \text{and} \quad \gamma_j = \begin{cases} \delta_j & , \quad j \le i-1 \\ 0 & , \quad j = i \\ \delta_{j-1} & , \quad j \ge i+1 \end{cases} .$$

Then the numerator is written analogous to (2.10) as

$$\exp(-\nu y_{n+1}) \sum_{k_{n+1}=0}^{\infty} \sum_{k_n=0}^{k_{n+1}} \cdots \sum_{k_1=0}^{k_2} \prod_{j=1}^{n+1} (\epsilon k_j)^{\gamma_j} \nu^{k_{n+1}} / \prod_{j=1}^{k_{n+1}} (k_j - k_{j-1})!$$

$$\times \prod_{j=1}^{n+1} (y_j - y_{j-1})^{k_j - k_{j-1}} \exp[\epsilon \sum_{j=1}^{n+1} ((n-j+1)y_j^0 - \sum_{m=j}^{n+1} y_m)(k_j - k_{j-1})] ,$$

$$x_{i-1} < t \le x_i , \quad i = 1, \cdots, n+1 . \tag{4.3}$$

Therefore, for right-censored samples the Bayes nonparametric estimator $\hat{\bar{F}}(t)$ of $\bar{F}(t)$ is given by (4.3) divided by (4.2). The numerator for the Bayes estimate of $r(t)$ can be found in a similar manner and is omitted here.

Remark 1. We remark that the Bayesian nonparametric estimator for $\bar{F}(t)$ given here is a smooth curve (and assumes F is an IFR distribution). Susarla and Van Ryzin's (1976) Bayesian nonparametric estimator for this case using the Dirichlet priors is discontinuous at the uncensored data points while the well-known Kaplan and Meier's (1958) estimator is a step function with jumps at the uncensored observations.

Remark 2. It should also be noted that we can obtain a Bayesian nonparametric estimator of the density function $f(t)$ in this framework. For the censored sample (X_i, δ_i), $i = 1, \ldots, n$, we calculate $f_n(t) = E[f(t) \mid (x_i, \delta_i), i = 1, \ldots, n]$ in a manner similar to that for (4.1). Since $f(t) = R(t) \bar{F}(t)$, the required numerator for $f_n(t)$ is the same as (4.3) with the stated relabeling of x_i's and

t to y_j's and δ_i's to γ_j's except that $\gamma_j = 1$ for $j = i$. Again, $\hat{f}_n(t)$ is a smooth density estimator and, of course, for $\delta_i = 1$ for all i, we have the case of no censoring as considered in Section 2.

ACKNOWLEDGEMENTS

REFERENCES

Antelman, G. and Savage, I. R.   (1965), Characteristic functions of stochastic integrals and reliability theory, Naval Res. Logist. Quart. 12, 199-222.

Barlow, R. E. and van Zwet, W.   (1971), Comparison of several nonparametric estimators of the failure rate function, Operations Res. and Reliability (Crouchko, ed.), 375-99.

Doksum, Kjell   (1974), Tailfree and neutral random probabilities and their posterior distributions, Ann. Prob. 2, 183-201.

Dubey, Satya D.   (1967), Some percentile estimators for Weibull parameters, Technometrics 9, 119-29.

Ferguson, Thomas S.   (1973), A Bayesian analysis of some nonparametric problems, Ann. Statist. 1, 209-30.

Ferguson, Thomas S.   (1974), Prior distributions on spaces of probability measures, Ann. Statist. 2, 615-29.

Ferguson, Thomas S. and Phadia, Eswar G.   (1979), Bayesian nonparametric estimation based on censored data, Ann. Statist. 7, 163-86.

Gaver, D. P.   (1963), Random hazard in reliability problems, Technometrics 5, 211-26.

Grenander, Ulf   (1956), On the theory of mortality measurement, Part II, Skan. Aktuarietidskr. 39, 125-53.

Kaplan, E. L. and Meier, P.   (1958), Nonparametric estimation from incomplete observations, J. Amer. Statist. Assoc. 53, 457-81.

Lagakos, S. W.   (1979), General right censoring and its impact on the analysis of survival data, Biometrics 35, 139-56.

Lo, Albert Y.   (1979), Bayesian nonparametric method for rate functions, (To appear - Dept. of Math. and Statist., Univ. Pittsburgh, Report.)

Marshall, Albert W. and Proschan, Frank   (1965), Maximum likelihood estimation for distributions with monotone failure rate, Ann. Math. Statist. 36, 69-77.

Padgett, W. J. and Wei, L. J.   (1979), Maximum likelihood estimation of a distribution function with monotone failure rate based on censored observations, (To appear).

Reynolds, David S. and Savage, I. Richard   (1971), Random wear models in reliability theory, Adv. Appl. Prob. 3, 229-48.

Susarla, V. and Van Ryzin, J.   (1976), Nonparametric Bayesian estimation of survival curves from incomplete observations, J. Amer. Statist. Assoc. 71, 897-902.

Watson, G. S. and Leadbetter, M. R. (1964a), Hazard analysis I, Biometrika 51, 175-84.

Watson, G. S. and Leadbetter, M. R. (1964b), Hazard analysis II, Sankhyā Ser. A 26, 101-16.

REPORT DOCUMENTATION PAGE

1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
AFOSR TR–79–1033		

4. TITLE (and Subtitle)
A BAYESIAN NONPARAMETRIC ESTIMATOR OF SURVIVAL PROBABILITY ASSUMING INCREASING FAILURE RATE.

5. TYPE OF REPORT & PERIOD COVERED
Interim rept.

6. PERFORMING ORG. REPORT NUMBER

7. AUTHOR(s)
W. J. Padgett and L. J. Wei

8. CONTRACT OR GRANT NUMBER(s)
F49620–79–C–0140

9. PERFORMING ORGANIZATION NAME AND ADDRESS
University of South Carolina
Department of Mathematics, Computer Science, &
Statistics   Columbia, S.C.   29208

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
61102F 2304/A5

11. CONTROLLING OFFICE NAME AND ADDRESS
Air Force Office of Scientific Research/NM
Bolling AFB, Washington, D.C.   20332

12. REPORT DATE
August 1979

13. NUMBER OF PAGES
18

14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)

15. SECURITY CLASS. (of this report)
UNCLASSIFIED

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)
Approved for public release; distribution unlimited

TR-42

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)
Bayesian nonparametric estimation; Jump process priors; Life testing; Failure rate function; Right censorship; Reliability.

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Bayesian nonparametric estimators of the survival function, the failure rate function, and the density function are obtained using jump processes as prior distributions on the space of increasing failure rate functions. The jump processes are more appealing intuitively than previously used Dirichlet processes, integral gamma processes, and processes neutral to the right and have a meaningful physical interpretation. Examples are given and the estimates are compared with the maximum likelihood estimates. In addition, the Bayesian nonparametric estimators are presented for arbitrarily right-censored observations.

DD $_{1\ JAN\ 73}^{FORM}$ 1473   EDITION OF 1 NOV 65 IS OBSOLETE